J. Phycol. **42,** 336–349 (2006)

DOI: 10.1111/j.1529-8817.2006.00209.x

MODELING PHYTOPLANKTON ABUNDANCE IN SAGINAW BAY, LAKE HURON: USING ARTIFICIAL NEURAL NETWORKS TO DISCERN FUNCTIONAL INFLUENCE OF ENVIRONMENTAL VARIABLES AND RELEVANCE TO A GREAT LAKES OBSERVING SYSTEM[1]

David F. Millie[2]

Florida Institute of Oceanography, University of South Florida and Florida Wildlife Research Institute, Florida Fish and Wildlife Conservation Commission, Saint Petersburg, Florida 33701, USA

Gary R. Weckman

Department of Industrial and Manufacturing Systems Engineering, Ohio University, Athens, Ohio 45701 USA

Ryan J. Pigg

College of Marine Science, University of South Florida and Florida Wildlife Research Institute, Florida Fish and Wildlife Conservation Commission, Saint Petersburg, Florida 33701, USA

Patricia A. Tester, Julianne Dyble[3]*, R. Wayne Litaker*

Center for Coastal Fisheries and Habitat Research, National Oceanic and Atmospheric Administration, Beaufort, North Carolina 28516, USA

Hunter J. Carrick

School of Forest Resources, Fisheries and Wildlife, Pennsylvania State University, University Park, Pennsylvania 16802, USA

and

Gary L. Fahnenstiel

Great Lakes Environmental Research Laboratory—Lake Michigan Field Station, National Oceanic and Atmospheric Administration, Muskegon, Michigan 49441, USA

Abstract Phytoplankton abundance, as chl *a*, in Saginaw Bay, Lake Huron was modeled using artificial neural networks. Suites of abiotic variables served as predictors for the trends/patterns in chl *a* concentrations. Spatial and temporal gradients of sampling stations throughout the bay were evident, with physical/chemical differences arising from hydrological/meteorological forcing and zebra mussel recruitment. Chlorophyll *a* concentrations displayed corresponding disparities; concentrations differed between the inner and outer bays and varied intra- and inter-annually. Trained networks reproduced the intrinsic variance and magnitude of chl *a* dynamics. Modeled-measured concentrations best approximated a 1:1 relationship in a hybrid network incorporating both supervised and unsupervised training whereas concentrations greater than 15 µg/L were underestimated in networks utilizing only supervised training, likely because of inadequate training data. Variables indicative of phytoplankton nutrition, acting as proxy measurements of algal biomass, and/or corresponding to descriptors of hydrological and meteorological forcing had the greatest influence upon modeled concentrations. A conjunctive decision tree and a novel sensitivity analysis provided rule-based information and comprehensible interpretation of relationships among multiple predictor variables. From this, the "knowledge" embedded in trained networks proved extractable and usable for ecological theory generation and/or decision making within water-quality problem solving. Forecasting initiatives within the developing Great Lakes Observing System may be best served by embedding neural networks in mechanistic models to quantitatively initialize variables, qualitatively delineate conditions for projecting ecological structure, and/or estimate deviations from predictability within mechanistic simulations.

Key index words: **algae; chl; ecosystem modeling; Laurentian Great Lakes; phytoplankton**

Abbreviations: **ANN, artificial neural network; Cl^-, chloride; DOC, dissolved organic carbon; K_d, light**

[1]Received 5 August 2005. Accepted 23 December 2005.
[2]Author for correspondence: e-mail david.millie@myfwc.com.
[3]Present address: Great Lakes Environmental Research Laboratory, National Oceanic and Atmospheric Administration, Ann Arbor, MI 48105, USA.

attenuation; MSE, mean square error; NO_X^--N, nitrite (NO_2^-) + nitrate (NO_3^-)-nitrogen; NH_4^+-N, ammonia-nitrogen; NOAA, National Oceanic and Atmospheric Administration; NID, network interpretation diagram; PCA, principal component analysis; PE, processing element; POC, particulate organic carbon; $PSiO_2$, particulate silica; orthophosphate-phosphorus, PO_4^{-3}-P Secchi, secchi depth; SiO_2, silica; Temp, temperature; TP, total phosphorus; TSS, total suspended solids

The increased incidence of natural disturbances, anthropogenic stressors, and/or human health risks throughout aquatic systems has necessitated accurate characterization of system-level stability and/or sustainability. Phytoplankton are the major source of (pelagic) carbon and have a fundamental influence on global climatology. Their biomass, compositional, and/or cellular responses to environmental variations are rapid and often predictable (Harris 1986, Frouin and Iacobellis 2002, Tester et al. 2003). As such, phytoplankton serve as "ecological sentinels." The ability to functionally link and conceptually model phytoplankton structure and function with/from system-level patterns and processes is central to resource stewardship and proactive mitigation of disturbances, stressors, and health risks.

An intuitive premise for ecological modeling is that interacting abiotic and biotic variables create environmental gradients responsible for the transitory and spatially explicit patterns of natural populations; as such, these variables may serve as predictors for phytoplankton dynamics and distributions (Hedgepeth 1977, Scavia et al. 1981, Bierman et al. 2005, Millie et al. 2006). However, predictable (linear) associations among environmental variables often do not exist and, as a consequence, modeling populations often is dependent upon characterizing complex non-linear, and stochastic interactions among patterns and processes (see Sugihara et al. 1990, Harris 1994, Mazumder 1994, Strutton et al. 1996, Jeong et al. 2003). In this regard, artificial neural networks (ANNs) show great promise for modeling phytoplankton abundance and dynamics (Recknagel 2003a). Neural networks are combinations of simple, non-linear models based on a "learn-by-example" paradigm (Huang and Xing 2002), whereby correlated patterns between input data (e.g. environmental conditions) and corresponding output values (e.g. phytoplankton biomass) are identified and reproduced. The "learned" models then are used to predict output values from new, independent data sets. However, deconvolving the interaction among predictor variables and delineating the influence of a single (or suites of) variables(s) on a desired output is difficult and often nonsensical (Garson 1991, Huang and Xing 2002, Olden and Jackson 2002, Millie et al. 2006).

Saginaw Bay, Lake Huron (Laurentian Great Lakes) has a long history of anthropogenic eutrophication and perturbation. Before 1980, water-column chl *a* and total phosphorus (TP) concentrations regularly exceeded 20 μg/L and 1 μM, respectively, with cyanobacteria typically dominating phytoplankton assemblages during hot, summer months. Concentrations of chl *a* and TP (along with corresponding phytoplankton blooms) decreased during the late 1980s and early 1990s, after aggressive nutrient abatement programs were initiated (Freedman 1974, Bierman et al. 1984, Makarewicz and Bertram 1991, Brittain et al. 2000). However, the reoccurrence of cyanobacterial blooms since 1994 (attributed to the long-term effects of nonindigenous zebra mussels, *Dreissena polymorpha* Pallas, Vanderploeg et al. 2001), present serious, recognized impacts to aquatic resources, human health, and community economics.

Identifying interactive quantifiers/predictors, environmental forcing factors, and data-assimilative modeling approaches for the forecasting of water quality and phytoplankton abundance within Saginaw Bay and throughout the Great Lakes is critical. Water column chl *a* concentration is a quantifier for (total) phytoplankton biomass and used to monitor system-level response to changing environmental variables (Paerl 1988, Millie et al. 1993, LaBaugh 1995, Paerl et al. 2003). Here, we (1) identified spatial/temporal patterns of diverse water-quality variables throughout Saginaw Bay; (2) using candidate abiotic variables, developed and validated ANNs for modeling chl *a* concentrations; (3) extracted, from trained networks, the impact and/or functional relevance of physical and chemical variables to phytoplankton abundance; and (4) discussed the appropriateness of incorporating statistical-based modeling into ecological forecasting initiatives within an evolving Great Lakes Observing System (GLOS).

MATERIALS AND METHODS

Study site and data acquisition. Saginaw Bay (Fig. 1), located along the eastern central shoreline of Michigan's lower peninsula, is a shallow, productive extension (approximately 2960 km^2) of Lake Huron, the third largest of the Great Lakes by volume (Fuller et al. 1995). The bay is subjectively divided into inner and outer bays, each having a distinct topography and physical/chemical characteristics (Table 1). The shallow inner bay (mean depth, 5.1 m) is mostly influenced by the Saginaw River, which accounts for 70% of total tributary inflows. The deeper outer bay (mean depth, 13.7 m) is heavily influenced by the cold, nutrient-poor waters of Lake Huron. Circulation is weak (7–11 cm/s) with water exchanges between Lake Huron and the outer bay and the inner and outer bays occurring mostly along the northern shore, with alterations attributable to wind events (Johengen et al. 1995, Nalepa et al. 1996). The bay's drainage basin encompasses approximately 22,556 km^2 of agricultural, industrial, and urban lands; excessive nutrient inputs (nearly two metric tons TP per day) from the basin during the 1970s and 1980s altered waters from a mesotrophic to eutrophic state (Fuller et al. 1995), with the inner bay being the most affected (Canale and Squire 1976, Dolan et al. 1978). Based on impairments of resource utilization arising from eutrophication, toxic substance and bacterial contami-

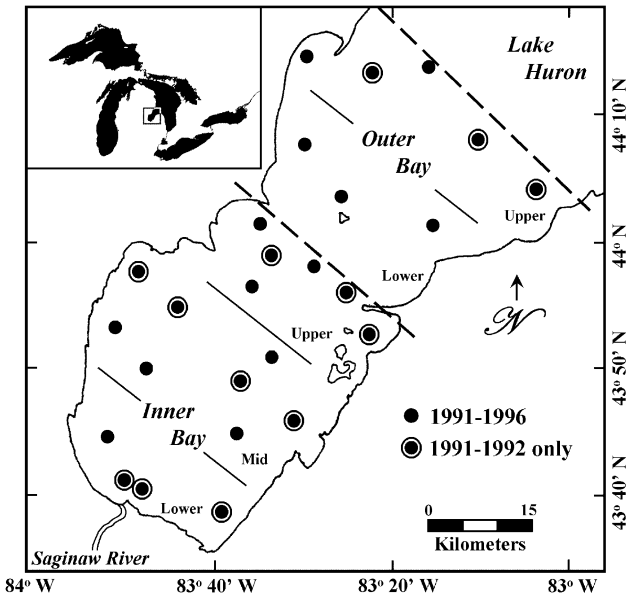

FIG. 1. Location of sampling stations within Saginaw Bay, 1991–1996. Inset figure places study area relative to Laurentian Great Lakes (USA).

nation, and sedimentation, Saginaw Bay was designated one of 43 "areas of concern" throughout the Great Lakes by the *Great Lakes Water Quality Agreement* (of 1972, amended in 1978) between the United States and Canada.

As part of a National Oceanic and Atmospheric Administration (NOAA) program assessing the impact of zebra mussels on Saginaw Bay structure and function, water quality data (Table 1) were acquired on a tri-weekly to monthly basis from April through November, 1991–1996. Data used herein and information concerning sampling sites and dates, collection methodology, sample processing, and analytical procedures were presented within NOAA Technical Reports, TM-091 (Nalepa et al. 1996) and TM-115 (Johengen et al. 2000, available at http://www.glerl.noaa.gov/pubs/techrept.html). Data summaries and interpretation, in part, were contained in papers published as a special issue to the *Journal of Great Lakes Research* (Nalepa and Fahnenstiel 1995).

TABLE 1. Subsurface water quality variables within the inner and outer Saginaw Bay, 1991–1996 (refer to Fig. 1).

Variable (units)	Inner bay ($I = 480$)	Outer bay ($n = 240$)
Temp (° C)*	16.46 ± 0.28	14.35 ± 0.39
Secchi (m)*	1.91 ± 0.05	5.03 ± 0.14
K_d (m^{-1})*	0.99 ± 0.04	0.35 ± 0.01
TSS (mg/L)*	7.57 ± 0.39	1.79 ± 0.11
TP (μg/L)*	18.40 ± 0.59	5.78 ± 0.22
PO_4^{-3}-P (μg/L)*	1.25 ± 0.11	0.78 ± 0.04
NO_X^--N (mg/L)*	0.41 ± 0.02	0.30 ± 0.01
NH_4^+-N (μg/L)	22.70 ± 1.25	18.88 ± 1.22
SiO_2 (mg/L)*	1.32 ± 0.05	1.03 ± 0.03
$PSiO_2$ (mg/L)*	1.05 ± 0.04	0.46 ± 0.03
Cl (mg/L)*	17.49 ± 0.31	8.75 ± 0.19
POC (mg/L)*	1.21 ± 0.04	0.44 ± 0.02
DOC (mg/L)*	3.61 ± 0.07	2.29 ± 0.03
chl *a* (μg/L)*	6.66 ± 0.29	2.35 ± 0.16

Data are means ± 1 SE.

*Indicates difference ($P \leq 0.05$) between inner and outer bays.

Temp, Temperature; K_d, light attenuation Secchi, secchi depth; TSS, total suspended solids; TP, total phosphorus; PO_4^{-3}-P, orthophosphate-phosphorus; NO_X^--N, nitrite (NO_2^-) + nitrate (NO_3^-)-nitrogen; NH_4^+-N, amonia-nitrogen; SiO_2, silica; $PSiO_2$, particulate silica; Cl^-, chloride; POC, particulate organic carbon; DOC, dissolved organic carbon; chl *a*, chlorophyll *a*.

Statistical analysis. To ascertain the symmetric distribution of and/or linear relationships among variables (and thereby determine a variable's adequacy for inclusion or elimination within statistical model development), the associations between physical, chemical, and biological variables within Saginaw Bay first were identified. The correspondence between variables was determined using Pearson product moment correlation coefficients (SYSTAT 10, 2000). Data were square root- or logarithmic-transformed (where appropriate) to stabilize variance and increase homogeneity of normalcy (Snedecor and Cochran 1980). Sampling sites and dates were characterized, with respect to physical and chemical variables, with a principal component analysis (PCA) utilizing euclidean distances (Clarke and Gorley 2001, Clarke and Warwick 2001). For the PCA, monthly means of variables were utilized. An unpaired *T*-test or an analysis of varaince (SYSTAT 10, 2000) assessed differences among water-quality variables between the inner and outer bays and spatial/temporal differences among monthly means of chl *a* concentrations.

Artificial neural networks. Concentrations of chl *a* were predicted from physical and chemical variables using ANNs incorporating both supervised training and a hybrid integration of supervised and unsupervised training (Fig. 2). For supervised training (in which values for input and output variables are known), multi-layer perceptrons using a backpropagation learning algorithm were constructed with NeuroSolutions v4.32 or 5.01 software (NeuroDimension Inc.; Gainesville, FL, USA):

$$[\text{chl}\,a] = f\{W_{P_1,P_3}[f(W_{X_1,P_1} \cdot X_1 + W_{X_2,P_1} \cdot X_2 \ldots W_{X_i,P_1} \cdot X_i + \varepsilon_1)]\} + f\{W_{P_2,P_3}[f(W_{X_1,P_2} \cdot X_1 + W_{X_2,P_2} \cdot X_2 \ldots W_{X_i,P_2} \cdot X_i + \varepsilon_2)]\} + f\{W_{P_j,P_3}[f(W_{X_1,P_j} \cdot X_1 + W_{X_2,P_j} \cdot X_2 \ldots W_{X_i,P_j} \cdot X_i + \varepsilon_j)]\} \quad (1)$$

where $X_{1,2,\ \ldots\ i}$ are candidate water quality predictor variables, $P_{1,2,3,\ \ldots\ j}$ are PEs, and $W_{X1,2,\ \ldots\ i,P1,2,3,\ \ldots\ j}$ are scalar weights, and $\varepsilon_{1,2,\ \ldots\ j}$ is the error. Hybrid networks implemented unsupervised training (in which inputs were preprocessed to identify and extract sets of uncorrelated linear features) immediately before supervised training. As the unsupervised procedure, PCA-derived eigenvectors were extrapolated from the correlation matrix of the input data vector using Hebbian learning (Principe et al. 2000).

Network structure and operation (including data normalization/derivation) followed that presented by Millie et al. (2006; refer to Fig. 2). Distinct training, cross-validation and testing data sets were selected randomly. For training ANNs, 60% of data vectors were presented 1000–2000 times to the network with this process repeated 20–100 times. "Learning" and momentum rates and step sizes were allowed to vary during iterative training to accelerate "learning" and to ensure network convergence to the global minimum (Barciela et al. 1999, Olden 2000, Principe et al. 2000, Olden and Jackson 2002). The mean square error (MSE) also was computed for a "cross validation" subset (15% of data vectors; after Olden 2000, Olden and Jackson 2002, Gurbuz et al. 2003) to provide an unbiased estimation of network prediction and to ensure optimal network design. If the MSE within the training or cross-validation data sets fell below 0.01 or began to increase (i.e. an indication that the network began to memorize the data; see Karul et al. 2000, Gurbuz et al. 2003), training was

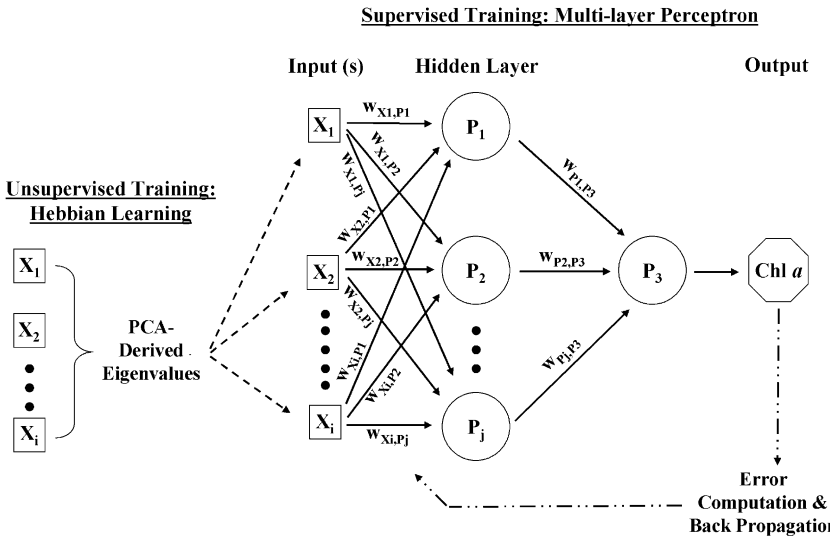

FIG. 2. Schematic of artificial neural network (ANN)s. Supervised training: a feed-forward, multi-layer perceptron with back-propagation depicting the interaction among water-quality data ("inputs"; $X_{1,\ldots i}$), processing element (PE)s ($P_{1,\ldots j}$), synaptic weights ($w_{X1,\ldots i,P1,\ldots j}$ and $w_{P1,\ldots j,P3}$), and the modeled chl *a* concentration ("output"). For each complete presentation of the data set (an epoch), input vectors were multiplied by scalar weights prior to processing by hyperbolic tangent functions (hidden layer), with subsequent values multiplied by weights and scaled via a linear-bias transfer function (output layer) to produce a modeled output. Modeled concentrations were "fed forward" for comparison to actual concentrations, from which the mean square error (MSE) was computed and "back-propagated", with the weights incrementally adjusted, through gradient descent, in the direction of the minimum error among PEs. As error minimized over multiple epochs, weight values stabilized and modeled concentrations increasingly approximated measured concentrations. Unsupervised training: optimal linear features were extracted (as eigenvectors via Hebbian learning) from water-quality vectors, after which the eigenvalues were used as input data for supervised training.

terminated. The trained ANN then was applied to a "testing" subset (25% of data vectors) not used in training and cross-validation.

A correlation coefficient measured the agreement between modeled and measured chl *a* concentrations. Linear regression determined "lines-of-best-fit" for modeled:measured relationships of chl *a* concentrations. An analysis of covariance tested whether slopes of regression estimations differed from the slope of a corresponding 1:1 modeled:measured relationship (SYSTAT 10, 2000).

Discerning predictor variable impacts and interaction. Several approaches were used to optimize the inclusion or exclusion of candidate predictor variables within networks and elucidate the qualitative and quantitative impacts of and relationships among predictor variables. Sensitivity analysis determined the variation in chl *a* concentration attributable to deviations of individual variables (while other variables remained fixed at their respective means; Principe et al. 2000); deviations from the mean within ± one and two standard deviations discerned the most influencing parameter(s) during common and disturbance conditions, respectively (Jeong et al. 2003). The most significant variables (i.e. those creating the greatest variation in chl *a*) subsequently were incorporated into new networks that were trained and tested, before final model selection. The magnitude/direction of synaptic weights within trained ANNs was depicted by a network interpretation diagram (NID) (Özesmi and Özesmi 1999, Olden 2000). The relative share of prediction associated with input variables then was determined from weight values using an algorithm originally proposed by Garson (1991), and later modified by Milne (1995) and Gedeon (1997).

Decision-tree classification and a novel sensitivity analysis were conducted to extract rule-based "knowledge" and illustrate relationships between/among predictor variables from the supervised network incorporating numerical and factor variables, respectively. The TREPAN algorithm constructed a symbolic, tree-based representation of predictive model outcomes (i.e. subjective groupings of chl *a* concentrations; <1.5, 1.51–2.5, 2.51–4, 4.1–9, 9.1–15, and >15 μg/L). Unlike traditional decision-tree structures, the TREPAN algorithm treats rule-extraction as an inductive learning problem and uses queries to induce a concise, accurate depiction that best approximates the predictive function of the trained network (Craven 1996, Craven and Shavlik 1996). Using an extension of traditional sensitivity analysis (where individual variables are deviated to denote alterations of the outcome variable), the interacting influence of multiple abiotic variables were assessed by varying two and more variables for each of the aforementioned chl *a* classifications (with the remaining variables kept constant at their respective means; Schleiter et al. 2003, G. R. Weckman et al., unpublished results).

RESULTS

With respect to physical and chemical parameters, surface waters at sampling sites within the inner and outer bays from 1991 to 1996 were distinct. Mean values for almost all water-quality variables were greater within the inner bay; however, the Secchi depth (Secchi) was greater within the optically clear waters of the outer bay than within the turbid waters of the inner bay (Table 1). Mean ammonia-nitrogen (NH_4^+-N) concentrations were similar between bays ($P>0.05$). The PCA included descriptors indicative of hydrological and meteorological forcing and together explained approximately 64% of the total variability; Secchi, light attenuation (K_d), and TP, particulate silica ($PSiO_2$), and particulate organic carbon (POC) concentrations and temperature (Temp) and nitrite (NO_2^-) + nitrate

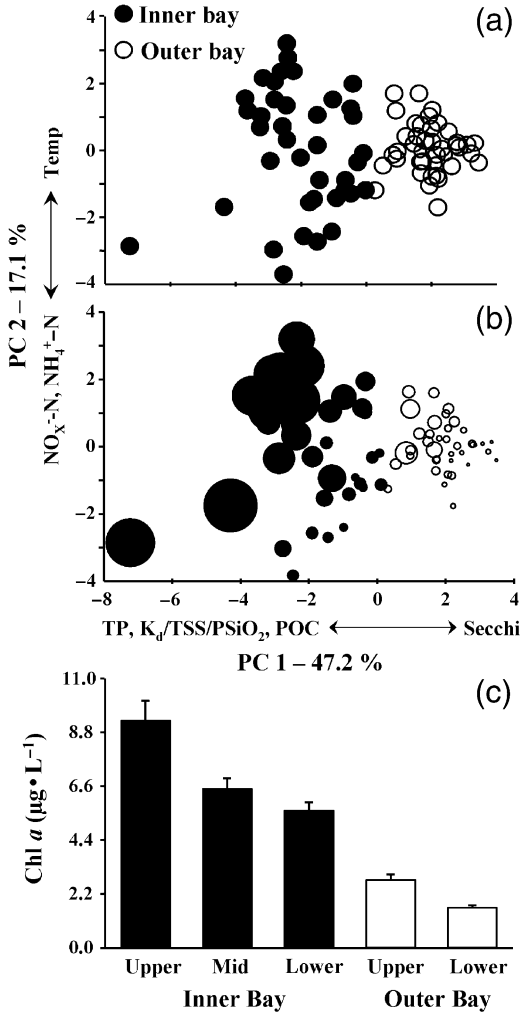

FIG. 3. (a and b) Two-dimensional ordination of sampling stations, as determined by a principal component analysis (PCA) of physical/chemical variables (see Methods). (a) Stations denoted as a function of inner or outer bay. (b) Mean chl *a* concentrations superimposed (as symbols of differing sizes—the larger the symbol, the greater the relative value) on ordination to identify spatial disparity in biomass. Numbers in parentheses along axes represent percent of total variability explained by the corresponding principal component (PC). (c) Concentrations of chl *a* as a function of bay reach (refer to Fig. 1), depicting concentration differences along a gradient from the Saginaw River to the lake proper. Data are means ± SE (n = 87–218).

(NO_3^-)-nitrogen (NO_X^--N), and NH_4^+-N concentrations explained approximately 47% and 17% of the variability within the first and second PC, respectively. From this, a spatial gradient of sampling sites from the inner to outer bay was evident (Fig. 3a). Phytoplankton abundance displayed a corresponding spatial disparity (Fig. 3b), with the mean chl *a* concentration for the inner bay greater than that for the outer bay. Concentrations maximized within the upper reach of the inner bay and decreased progressively to the lower reach of the outer bay ($P \leq 0.001$, Fig. 3c).

Temporal differences of physical/chemical variables also were apparent among sampling sites (although less so than spatial differences). Sites reflecting late summer/early fall sampling were dissimilar from sites reflecting spring/late fall sampling (Fig. 4a). Inter-annual variation among sites was less apparent; however, sampling sites during 1991 and 1992 were relatively distinct from sites during other years (Fig. 4c). Monthly means of chl *a* concentrations varied both intra- and inter-annually ($P = 0.002$, $F = 3.29$, df = 7 and $P < 0.001$, $F = 7.61$, df = 5), after allowing for differences between bays. Differences in mean concentrations primarily occurred within the inner bay, with the greatest concentrations during May, August, and September and the least during June and November (Fig. 4b). The greatest concentrations occurred during 1991 and 1994, with the least occurring during 1993 (Fig. 4d).

Supervised ANNs. Numerous networks, incorporating varied numbers of PEs and hidden layers, were developed to optimize prediction of chl *a* concentrations. Optimal models were selected, solely on training and testing performance (i.e. minimization of error, correspondence between predicted and measured concentrations, deviations of estimated regression lines from 1:1 modeled:measured relationships).

An ANN, utilizing 12 candidate physical and chemical variables (Temp, Secchi, K_d, chloride (Cl^-), NO_X^--N, NH_4^+-N, PO_4^{-3}-P, TP, silica (SiO_2), $PSiO_2$, dissolved organic carbon (DOC), POC) and comprised of one hidden layer with four PEs, provided the most reliable estimates of chl *a* concentrations. Values of MSE for both the training and cross-validation data subsets quickly approached zero (Fig. 5a), indicating that the network had succeeded in "training" the model. Upon applying the network to the testing data subset, modeled data mirrored the general trend in chl *a* dynamics ($\rho = 0.89$, $P \leq 0.0001$; Fig. 5b). A trained ANN utilizing 16 inputs (incorporating the aforementioned 12 inputs and four factor variables, denoting inner or outer bay, and surface or subsurface samples) provided similar prediction of chl *a* ($\rho = 0.90$, $P \leq 0.0001$; Fig. 5c). Sensitivity analyses for the original (12 input) network, in which variables were altered one and two standard deviations around their respective means (Figs. 6, a and b, respectively), denoted POC, $PSiO_2$, and TP to be the most important variables for predicting chl *a*. An ANN, utilizing only these three variables and comprised of one-hidden layer with 14 PEs, was successfully trained and cross validated (Fig. 6c) with no observed improvement (over the original 12 input model) in predicting chl *a* ($\rho = 0.89$, $P \leq 0.0001$; Fig. 6d). In all networks, however, modeled concentrations underestimated measured concentrations (to varying degrees) when concentrations exceeded approximately 15 µg/L (e.g. Figs. 5, b and c, and 6d). Although slopes of modeled:measured regressions differed from a corresponding 1:1 relationship for all supervised networks, the mean square and absolute errors (6.86 and 1.59, respectively) for the network incorporating factor variables (inner/outer bay and subsurface/depth

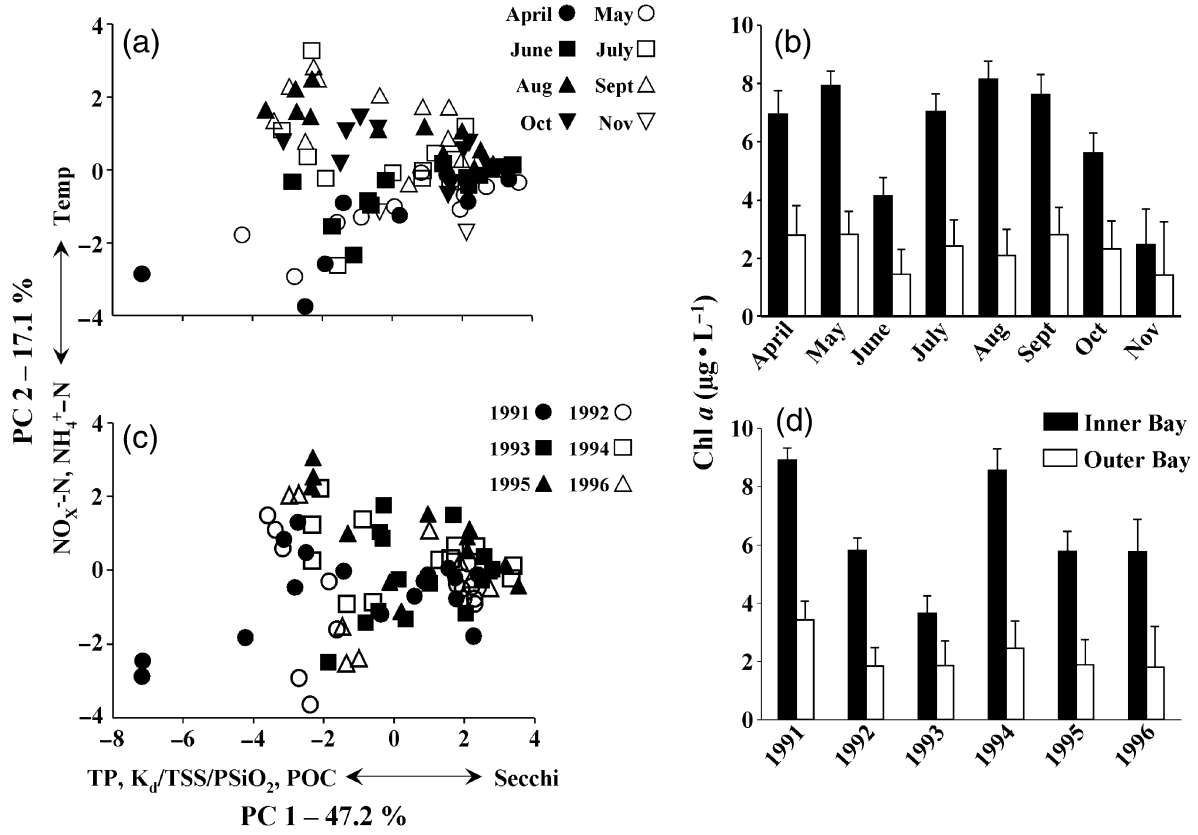

FIG. 4. Two-dimensional ordination of sampling stations, as determined by a principal component analysis (PCA) of physical/chemical variables. Stations denoted as a function of (a) sampling month and (c) year. Numbers in parentheses along axes represent percent of total variability explained by the corresponding principal component (PC). Concentrations of chl *a* within the inner and outer bays as a function of (b) sampling month (n = 8–100) and (d) year (n = 13–143). Data are means ± SE.

sample designation) was less than those for all other networks, indicating a slightly superior model. Attempts to optimize performance of this network through reduction of inputs by sensitivity analyses did not improve prediction (data not shown).

Hybrid ANNs. The dimensionality of the input data set was reduced to single linear combination (of five eigenvectors) for each data vector. The hybrid network, coupling this unsupervised training with supervised training utilizing four PEs within a hidden layer, provided similar predictive capacity ($\rho = 0.89$, $P \leq 0.001$; Fig. 7a) for chl *a* concentrations as that of supervised ANNs, but with lesser values for mean-square and mean-absolute error. Although not different, the slope of the regression for measured:modeled concentrations (0.80) more closely approximated a 1:1 relationship than those for supervised ANNs (0.56–0.69), indicting a more realistic prediction across the entire range of chl *a* concentrations. Sensitivity analyses (in which the initial input data were varied, before eigenvector derivation) identified Secchi, $PSiO_2$, POC, and Temp to have the greatest impact on network prediction (Fig. 7b).

Variable impact and interaction. Prediction of chl *a* in supervised ANNs (e.g. Fig. 8a) involved extremely complex interactions among physical and chemical variables. Few variables had consistent magnitude or direction (positive or negative), among all PEs within trained networks. The apparent strong positive influences of a few variables within some PEs generally were "counter-balanced" by equal or lesser negative influences of the same variables among alternative PEs. Based on absolute weight values from trained ANNs, Garson's algorithm did not denote any particular one (or few) variable(s) to greatly influence chl *a* (to the exclusion of other variables); rather multiple variables had similar relative effects (approximately 5% to 13%; Fig. 8b).

The TREPAN algorithm induced a conjunctive ("if-then") decision-tree that best predicted classifications of chl *a* throughout Saginaw Bay. Using variable queries for POC, TP, $PSiO_2$, SiO_2, Secchi, and NH_4^+-N, the decision tree delineated classifications across the entire range of chl concentrations, with prediction for chl *a* concentrations exceeding 4 µg/L generally distinct from that of lesser concentrations (Fig. 9). Multiple-variable sensitivity analysis illustrated the interacting impact of two- and three-variables on the predictive classifications of chl *a*. For example, although TP and Temp and TP, Temp, and Secchi had similar interact-

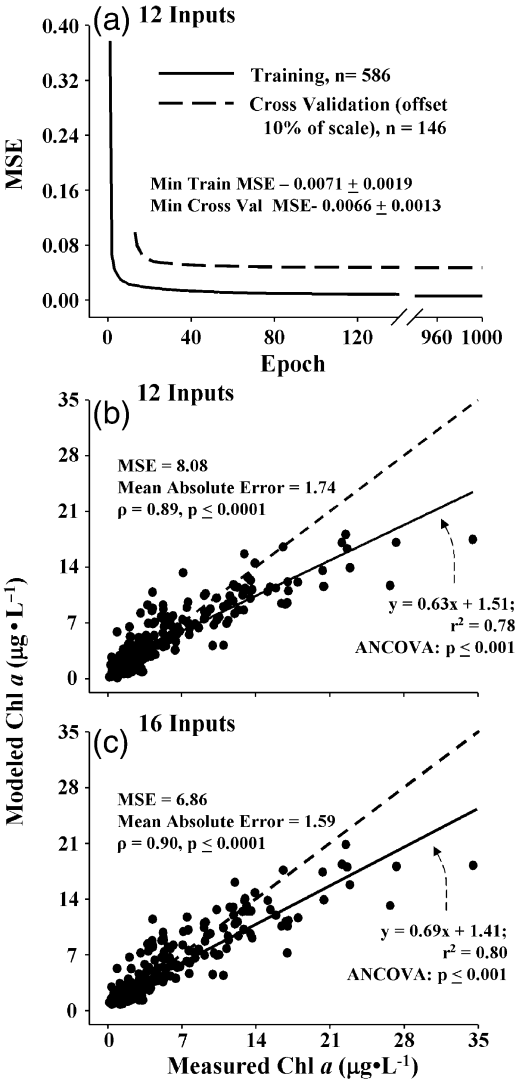

FIG. 5. (a) Mean square error (MSE) associated with training and cross-validation data during training of a supervised ANN incorporating 12 predictor variables (refer to Fig. 2). Data are means, $n = 100$. (b) Modeled chl *a* concentrations as a function of measured concentrations for the ANN incorporating 12 predictor variables ($n = 244$). (c) Modeled chl *a* concentrations as a function of measured concentrations for the ANN incorporating 16 predictor variables ($n = 244$). The solid line and corresponding statistical information represent the "best" fit for the modeled:measured relationship, as derived from linear regression. The dashed line represents a 1:1 relationship.

ing relationships throughout the entire bay (Figs. 10, a and c, b and d, respectively), the range and magnitude of the relationship within the inner and outer bays were distinct, particularly at lower temperatures.

DISCUSSION

Phytoplankton abundance throughout the Great Lakes reflects the balance between growth and loss (grazing and sedimentation), which varies seasonally (Fahnenstiel and Scavia 1987, Scavia and Fahnenstiel 1987, Carrick et al. 1991, 1992, Fahnenstiel et al. 2000, Carrick 2005). Within coastal waters, growth and accumulation further are mediated through episodic, short-term (hours to weeks) forcing phenomena such as storm-induced resuspension (Schelske et al. 1974, Millie et al. 2002, 2003), water-column advection (Fahnenstiel et al. 1988), and tributary inflows (Schelske et al. 1984, Klarer and Millie 1994, Lohrenz et al. 2004). Spatial and temporal gradients of sampling sites were evident within Saginaw Bay, with differences attributable to suites of physical and chemical variables arising from meteorological and hydrological forcing, and the impacts of zebra mussel invasion. Phytoplankton abundance, as chl *a*, mirrored physical/chemical gradients; concentrations were greatest in the lower reaches of the inner bay and progressively decreased lake-ward, reflecting the decreasing impact of nutrient-laden inflows with increasing distance from mouth of the Saginaw River (refer to Fig. 1). Within the inner bay, chl *a* concentrations were greatest during May and July–September, presumably reflecting a phytoplankton pulse with increased inflows, temperatures, and day length during late spring and the establishment of cyanobacterial assemblages upon the onset of high temperatures and a stagnated water column during mid- to late-summer, respectively (Vanderploeg et al. 2001, Bierman et al. 2005).

Supervised and hybrid ANN's characterized the complex non-linear, and stochastic interactions among diverse physical and chemical variables throughout the bay and together reproduced the intrinsic intra- and inter-annual variance (and magnitude) of chl *a* dynamics. As such, interacting suites of physical and chemical variables served as useful predictors for the trends/patterns of phytoplankton abundance. The (3–4 weeks) sampling resolution within the Saginaw Bay data set initially might not be expected to typify algal growth (that often is temporally "lagged"; Duarte 1990) and/or resolve episodic, short-term variability in meteorological and water-column conditions, thereby prohibiting the assertion of causality for phytoplankton accumulation. Lee et al. (2003) noted that a minimum sampling interval of one week was necessary for an ANN to accurately reproduce coastal phytoplankton dynamics. However, a back-propagation network (such as that used here)—if provided enough training data and PEs and given the identifiable function has a finite number of discontinuities—should serve as a universal approximator for delineating intra-annual/seasonal/cyclic trends among most (limnological) variables (Smith and Mason 1997, Maier et al. 1998, Karul et al. 2000).

Concentrations of chl *a* greater than 15 µg/L were underestimated in supervised networks, likely because of inadequate data representation within training data sets. Maximal phytoplankton biomass within Saginaw Bay occurred discontinuously, typically in late spring and/or late summer. Consequently, our sampling resolution provided a data set containing only few concentrations greater than 15 µg/L with which the ANNs

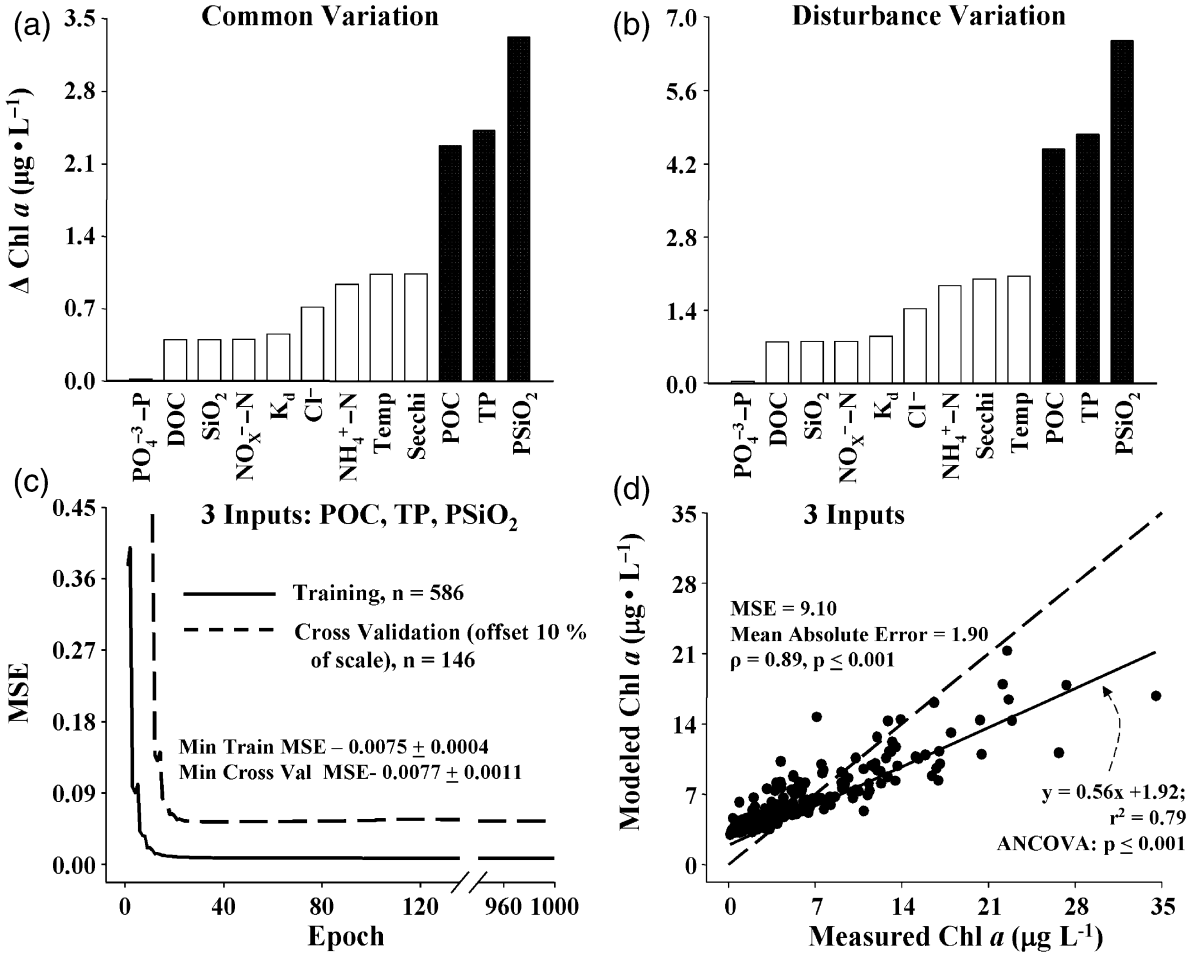

FIG. 6. Results of a sensitivity analysis across (a) common (±1 SD) and (b) disturbance (±2 SD) variations performed on training data from the supervised artificial neural network (ANN) incorporating 12 predictor variables. Black-filled bars indicate variables selected for subsequent modeling. (c) Mean square error (MSE) associated with training and cross-validation data subsets during training of the supervised ANN using the three variables identified by sensitivity analysis (refer to Fig. 2). Data are means, $n = 20$. (d) Modeled chl *a* concentrations using variables selected by a sensitivity analysis as a function of measured concentrations ($n = 244$). Lines and statistical information as in Fig. 5.

could be adequately trained (see Millie et al. 2006). To minimize this dilemma, hybrid ANNs reconstructed diverse water-quality variables onto a small dimensionality space (of eigenvectors). This data preprocessing maximally preserved variance within, while simultaneously extracting linear features from multiple suites of water quality variables. Although predictive capacity (in terms of ρ) was not greater for the hybrid network than that of supervised networks, a more optimal prediction across the entire range of "test" concentrations resulted (Bowden et al. 2003). Uncertainty exists as to whether an untried variation (number of hidden layers, PEs, learning rates, etc.) of a supervised ANN might have better extrapolated modeled values outside the range of the measured concentrations within the training set (Karul et al. 2000), thereby providing a predictive capability comparable with that of the hybrid network.

Concentrations of TP, POC, and $PSiO_2$ were identified to have the greatest influence upon modeled chl *a* concentrations (across both common and disturbance variations) within supervised networks. The inclusion (and importance) of such variables within networks for Saginaw Bay is intuitive; phosphorus is the nutrient considered most limiting for phytoplankton photosynthesis and growth throughout the Great Lakes (Schelske et al. 1974, 1986;, Schelske 1979, Hartig and Wallen 1984, Fahnenstiel et al. 1998) and variables acting as proxy measurements, to any significant degree, for phytoplankton abundance (i.e. POC and $PSiO_2$ for all algae and diatoms, respectively) would increase the predictive capability for chl *a* (Millie et al. 2006). Modeling the impact of nutrient perturbations on the stability of microbial food-web dynamics within central Lake Erie, Heath et al. (2000) also noted that phosphorus concentration was the major "forcing" factor in their simulations. Remarkably, the use of only three variables as predictors within supervised networks provided a prediction capability similar to that of supervised/hybrid networks using

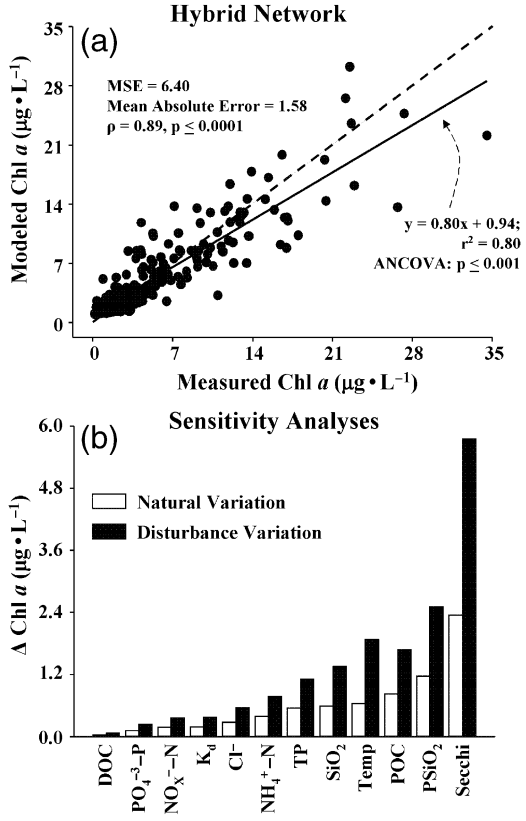

FIG. 7. (a) Modeled chl *a* concentrations as a function of measured concentrations for the hybrid artificial neural network (ANN) ($n = 244$). Lines and statistical information as in Fig. 5. (b) Results of a sensitivity analysis across common (± 1 SD) and disturbance (± 2 SD) variations performed on the training data subset from the hybrid ANN.

all abiotic variables. However, no one (or few) variable(s) greatly impacted chl *a* (on a relative basis), to the exclusion of others (Fig. 8b). Rather, multiple abiotic variables interacted to impact trends and patterns of phytoplankton abundance, signifying the holistic complexity of physical and chemical relationships throughout the bay. In addition, multiple variables (TP, $PSiO_2$, POC, Secchi, Temp) greatly influenced chl *a* concentrations within the hybrid network. This selection of variables is insightful; the chosen variables corresponded to (groups of) variables explaining the majority of variance associated with the first and second PCs (compare Fig. 8b with Figs. 3 and 4) and were descriptors indicative of hydrological and meteorological differences among sampling sites on an intra- and inter-annual basis.

Water clarity in the inner bay typically is influenced by River discharge, phytoplankton production, and wind-driven resuspension during the spring, summer, and fall, respectively (Budd et al. 2001). The recruitment of zebra mussels (beginning in fall of 1991) dramatically impacted limnological conditions; chl *a* and TP concentrations decreased with a corresponding increase in water clarity (Fahnenstiel et al. 1995, Nalepa et al. 1995, 1999, Budd et al. 2001) and nutrient utilization by phytoplankton appeared to slow as mussel excretion provided an alteration between phosphorus-limited and -replete conditions (Heath et al. 1995, Bierman et al. 2005). Undoubtedly, the inclusion of mussel density and/or biomass as (a) variable(s) would have altered (improved?) prediction of phytoplankton abundance, albeit ρ values for modeled:measured concentrations approximated 0.90 for all ANNs. It also may be argued that the contemporary water quality variables utilized were the result of and reflected the impact(s) of zebra mussels on the holistic physical and chemical characteristics of the water column. In any event, comprehensive spatial/temporal data concerning zebra mussel occurrence was not available for network inclusion; mussel data only was available at most, for only six stations and for one fall sampling date annually (Nalepa et al. 2002).

An important attribute for ecological modeling is to derive the importance of and relationships among abiotic (predictor) variables, thereby providing the bases for scientific theory generation and/or decision making within water-quality problem solving (e.g. Reckhow 1994). Here, networks were developed to optimize prediction of chl *a*. However, ANNs typically have high-dimensional input space with no explicit, declarative knowledge structure (Craven 1996, Huang and Xing 2002); the "knowledge" of trained networks is encoded (almost incomprehensibly) within the interacting complexity of synaptic weights among input-hidden-out layers (e.g. Fig. 7a). As such, it is not at all surprising that most ecologists consider ANNs to be "black boxes" and treat them as numerical enigmas (Olden and Jackson 2002). Clearly, qualitative and quantitative information concerning the environmental variables used to capture (and explain) the predictive outcomes need to be extracted from complex ANNs modeling the timing and magnitude of phytoplankton accumulation (Wigham and Fogel 2003, Recknagel 2003b).

Simple, visualization techniques (easier to interpret than NIDs) illustrating network prediction complexity and interaction are required to generalize individual and/or multiple abiotic influences upon chl *a* throughout dynamic coastal waters. To this end, the knowledge and "rule" sets embedded in an optimal ANN for Saginaw Bay were emulated and illustrated. The TREPAN algorithm provided variable translation for network predictive outcomes across high-dimensional space (i.e. 16 predictor variables) and induced a comprehensible, conjunctive decision matrix for predictive chl *a* classifications. Although the trained network from which the decision tree was induced incorporated both numerical and factor predictor variables, only numerical values delineated outcome classifications (further highlighting the distinct interacting physical and chemical relationships throughout the inner and outer bays and their holistic importance in intra- and inter-annual variation of chl *a*). Interestingly, the variables designated within tree "nodes" also were identified as "best predictor" variables by traditional

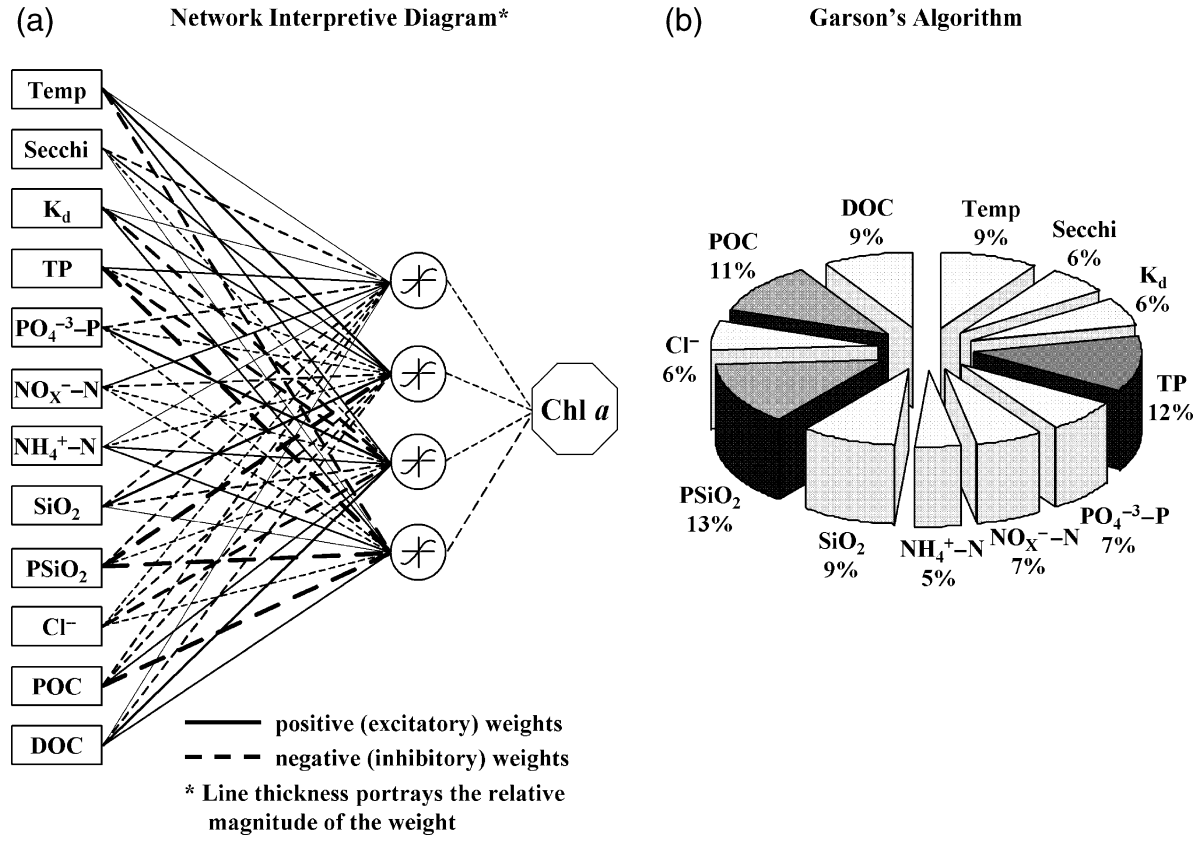

FIG. 8. (a) A network interpretation diagram (NID) for the supervised artificial neural network (ANN) incorporating 12 predictor variables. Dashed and solid lines depict negative (inhibitory) or positive (excitatory) effects, respectively, upon modeled concentrations by synaptic weights among input-hidden-output layers. Line thickness portrays the relative magnitude of the weight (i.e. greater values indicate more significance in prediction, compared to lesser values). Contrasting inhibitory/excitatory weights entering the same processing element (PE) identifies variable interaction (from Olden 2000, Olden and Jackson 2002). (b) The relative share of prediction associated with abiotic/biotic variables in modeling chl *a* concentrations, as determined using Garson's Algorithm. Dark shade denotes variables having the greatest impact on prediction (see Results).

sensitivity analyses for both supervised and hybrid networks (see above). However, such sensitivity analyses can only provide information concerning how network outcomes (e.g. chl *a*) may or may not be altered by deviation of a single variable (e.g. TP). Improving on this obvious limitation, multi-variable sensitivity analysis provided interpretable illustrations for interacting two-dimensional (e.g. TP, Temp) and three-dimensional (e.g. TP, Temp, Secchi) contributions between/among abiotic variables on predictive classifications of chl *a*. From such graphical representations, not only can interacting contributions of variables within distinct portions of the bay be identified, but the interacting impact or contribution of potentially "hidden" variables also may be delineated. In particular, distinct mathematical expressions approximating the linear/curvilinear relationship and/or interaction of multiple abiotic variables upon model outcomes can be (easily) generated for future model formulation.

Relevance of ANNs to a GLOS. The Great Lakes constitute the largest system of fresh, surface water on earth (containing approximately 18% of the world supply); as such, they are an invaluable natural resource (for consumption, transportation, power, recreation, etc.) and impact global climatology and carbon cycling (Fuller et al. 1995). Effective resource management for these waters requires timely and accurate assessment and prediction of biological outcomes in response to ecosystem change. However, before the routine assimilation of ecological modeling into Great Lakes management strategies, the means to integrate prediction with interpretable process-level information over dynamic spatial and temporal scales is required.

Coastal assessment programs typically have depended upon adaptive, invasive sampling with its intermittent data acquisition, resulting in discontinuous "snapshots" of the ecological continuum. These programs have become increasingly reliant upon autonomous data acquisition often integrated across multi-platform sampling networks and observatories, with their potential for real-time interpretation of system-level variability (Dickey 2003, Moersdorf and Meindl 2003, Schofield and Tivey 2004, Paerl et al. 2005). Microelectronic-based probes/sensors (biological, chemical, physical) and fixed/transportable sampling

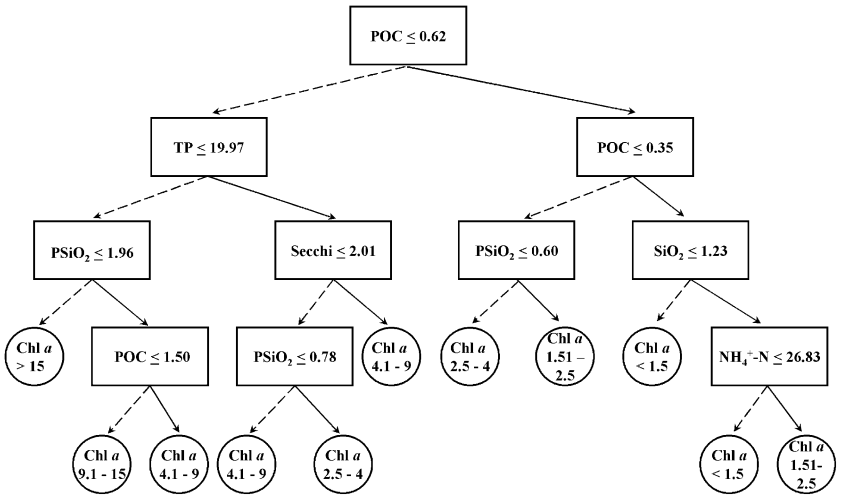

FIG. 9. A conjunctive classification tree, derived using the TREPAN algorithm (see Methods) for the supervised artificial neural network (ANN) incorporating 16 predictor variables. Refer to Table 1 for variable abbreviations and units. Dashed and solid lines depict negative or positive responses, respectively, to declared variable values/limits within decision nodes.

platforms are evolving technologies, providing a never-before recognized sampling resolution across ecological, system, and organismal scales. The resulting (and potential) increase in data acquisition necessitates a requirement for novel data manipulation/minimization, pattern recognition, and information synthesis.

Neural networks developed for Saginaw Bay utilized abiotic data collected through both invasive and instrumental-based sampling protocols. Direct quantification of select abiotic variables (Temp, PO_4^{-3}-P, NO_X^--N, NH_4^+-N, SiO_2, Cl^-) is possible using readily available autonomous probes and sensors. Other variables can be approximated, to varying degrees of success, from data acquired through nutrient probes (TP from PO_4^{-3}-P), PAR sensors (K_d), and transmissometers [total suspended solids (TSS), POC]. When these 10 abiotic variables were used as predictors in a distinct supervised network (comprised of one-hidden layer optimized with four PEs), the correspondence ($\rho = 0.87$; $P \leq 0.0001$) between measured and modeled chl *a* concentrations and predictive capacity of the network ($r^2 = 0.76$; $P \leq 0.001$) were comparable with that obtained in previous supervised/hybrid networks using all available variables. Concentrations of POC, TP, and NH_4^+-N were identified (via sensitivity analysis) to have the greatest influence on modeled chl *a* concentrations in this ANN, consistent with previous findings denoting these variables to discern (in part) differences among sampling sites, greatly impact prediction in ANNs, and act as discriminating "nodes" for predictive chl *a* classifications in the conjunctive decision matrix. From this, it appears that chl *a* concentrations throughout the physically and chemically distinct inner and outer bays may be dependably modeled using abiotic data acquired via contemporary autonomous technology.

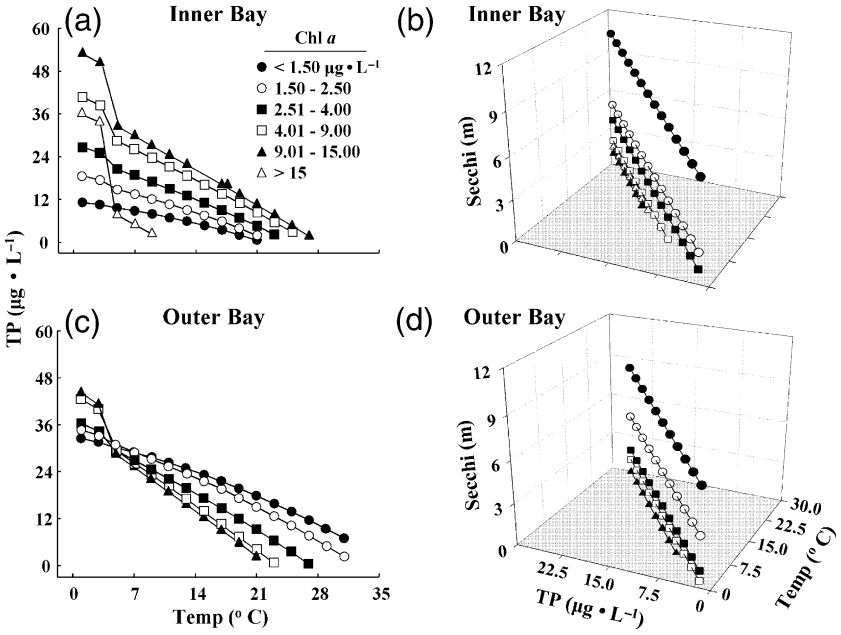

FIG. 10. Results of a multi-variable sensitivity analysis performed on training data from the supervised artificial neural network (ANN) incorporating 16 predictor variables. (a and c) Relationships between temperature (Temp) and total phosphorus (TP) across classifications of chl *a* concentration within the inner and outer bays, respectively. (b and d) Relationships between Temp, TP, and Secchi depth (Secchi) across classifications of chl *a* concentration within the inner and outer bays, respectively.

Such network-based prediction of chl *a* can assist data synthesis and knowledge-based modeling within forecasting initiatives for a developing GLOS (http://www.glc.org/glos/; Schwab et al. 1992, Schwab and Bedford 1999). Adaptive models addressing the timing and proliferation of toxic cyanobacterial abundance in response to nutrient, meteorological, and hydrological forcing, along with potential human-health impacts, are a desirable component for the GLOS. Mechanistic modeling of Great Lakes phytoplankton and associated abiotic/biotic forcing functions (Scavia et al. 1981, Heath et al. 2000, Bierman et al. 2005) may be best served by a modular framework in which ANNs (with corresponding' rule-extraction' methodologies) are embedded within system-level, mechanistic models and used to quantitatively initialize variables, qualitatively delineate conditions for projecting ecological structure; and/or estimate deviations from predictability within mechanistic simulations (Murray and Parslow 1999).

For example, Walsh et al. (2001, 2003) utilized a framework of ecological models embedded in a physical model to simulate the onset of toxic phytoplankton blooms in response to diverse nutrient regimes throughout Florida coastal waters. However, their models required enormous sampling-derived data sets for parameter estimation, specification of the initial phytoplankton biomass, and quantification of abiotic/biotic interactions known to affect bloom dynamics and rate processes. Ambiguity in the "initialization" step, omission of pertinent parameters, and/or inclusion of parameters with uncertain "boundary condition" can result in simulation forecasting that is at best, only "crude approximations of reality" (after Maier et al. 1998). Initialization parameters and delineation of variable "boundary conditions" for Great Lakes cyanobacterial assemblages could be continuously derived/altered via ANNs embedded within GLOS forecast simulations and incorporating historical abiotic/biotic data, with networks "updated" as new data become available.

Knowledge gained from a modular modeling framework would allow the integration of the interactive physical, chemical, and biological forces that drive ecological change with concurrent estimation of phytoplankton abundance along physical–chemical gradients and ecosystem to regional scales. Moreover, when combined with event–response confirmation (http://www.glerl.noaa.gov/res/Centers/HumanHealth/hab/EventResponse/EventResponse.html), an effective cyanobacterial model would assist in providing an objective early warning system for alerting scientists, water resource managers, fisheries personnel, and public health officials to instances/impacts of toxic bloom occurrence (Millie et al. 1999). Products from such a multi-faceted approach then can be assimilated into ecological data portals (e.g. AIBS/NSF's *National Environmental Observatory Network*, NOAA's *Integrated Ocean Observing System*, and *Ocean and Human Health Initiatives*) facilitating regional-scale data dissemination efforts and subsequently providing for the development/application of scientifically based environmental and public-health policy.

This work is a portion of the multi-disciplinary research program of the Center of Excellence for Great Lakes and Human Health, funded by the Oceans and Human Health Initiative of NOAA's Office of Global Programs and the Cooperative Institute for Limnology and Ecosystems Research, a joint institute of the University of Michigan and NOAA's Office of Oceanic and Atmospheric Research. Reference to proprietary names are necessary to report factually on available data; however, the Florida Institute of Oceanography, Ohio University, NOAA, Pennsylvania State University, University of South Florida, Florida Fish and Wildlife Conservation Commission, and University of Michigan neither guarantee nor warrant the standard of a product and imply no approval of a product to the exclusion of others that may be suitable.

Barciela, R. M., Garcia, E. & Fernández, E. 1999. Modelling primary production in a coastal embayment affected by upwelling using ecosystem models and artificial neural networks. *Ecol. Model.* 120:199–211.

Bierman, V. J., Dolan, D. M., Kasprzyk, R. & Clark, J. L. 1984. Retrospective analysis of the response of Saginaw Bay, Lake Huron, to reductions in phosphorus loadings. *Environ. Sci. Technol.* 18:23–31.

Bierman Jr., V. J., Kaur, J., DePinto, J. V., Feist, T. & Dilks, D. 2005. Modeling the role of zebra mussels in the proliferation of blue-green algae in Sagniaw Bay, Lake Huron. *J. Great Lakes Res.* 31:32–55.

Bowden, G. J., Dandy, G. C. & Maier, H. R. 2003. An evaluation of methods for the selection of inputs for an artificial neural network based river model. *In* Recknagel, F. [Ed.] *Ecological Informatics; Understanding Ecology by Biologically-Inspired Computation*. Springer-Verlag, Berlin, pp. 215–32.

Brittain, S. M., Wang, J., Jackson, L. B., Carmichael, W. W., Rinehart, K. L. & Culver, D. A. 2000. Isolation and characterization of microcystins, cyclic heptapeptide hepatotoxins from a Lake Erie strain of *Microcystis aeruginosa*. *J. Great Lakes Res.* 26: 241–49.

Budd, J. W., Drummer, T. D., Nalepa, T. F. & Fahnenstiel, G. L. 2001. Remote sensing of biotic effects: zebra mussels (*Dreissena polymorpha*) influence on water clarity in Saginaw Bay, Lake Huron. *Limnol. Oceanogr.* 46:213–23.

Canale, R. P. & Squire, J. 1976. A model for total phosphorus in Saginaw Bay. *J. Great Lakes Res.* 2:364–73.

Carrick, H. J. 2005. An under-appreciated component of biodiversity in plankton communities: the role of protozoa in Lake Michigan (a case study). *Hydrobiologia* 551:17–32.

Carrick, H. J., Fahnenstiel, G. L., Stoermer, E. F. & Wetzel, R. G. 1991. The importance of zooplankton–protozoan trophic couplings in Lake Michigan. *Limnol. Oceanogr.* 36:1335–45.

Carrick, H. J., Fahnenstiel, G. L. & Taylor, W. D. 1992. Growth and production of planktonic protozoa in Lake Michigan: in situ versus in vitro comparisons and importance to food web dynamics. *Limnol. Oceanogr.* 37:1221–35.

Clarke, K. R. & Gorley, R. N. 2001. *PRIMER v5: User Manual/Tutorial*. Primer-E, Plymouth, UK, 91 pp.

Clarke, K. R. & Warwick, R. M. 2001. *Change in Marine Communities: An Approach to Statistical Analyses and Interpretation*. 2nd ed. Primer-E, Plymouth, UK.

Craven, M. W. 1996. *Extracting comprehensible models from trained neural networks*. Ph.D. dissertaion, University of Wisconsin-Madison. 199 pp.

Craven, M. W. & Shavlik, J. W. 1996. Extracting tree-structured representations of trained networks. *In* Touretzky, D., Mozer, M. & Hasselmo, M. [Eds.] *Advances in Neural Information Processing Systems*. 8. Massachusetts Institute of Technology Press, Cambridge, pp. 24–30.

Dickey, T. 2003. Emerging ocean observations for interdisciplinary data assimilation systems. *J. Mar. Syst.* 40–41:5–48.

Dolan, D. M., Bierman Jr., V. J., Dipert, M. H. & Geist, R. D. 1978. Statistical analysis of the spatial and temporal variability of the ratio of chl *a* to phytoplankton cell volume in Saginaw Bay, Lake Huron. *J. Great Lakes Res.* 3:65–71.

Duarte, C. M. 1990. Time lags in algal growth: generality, causes and consequences. *J. Plankton Res.* 12:873–83.

Fahnenstiel, G. L., Bridgeman, T. B., Lang, G. A., McCormick, M. J. & Nalepa, T. F. 1995. Phytoplankton productivity in Saginaw Bay, Lake Huron: Effects of zebra mussel (*Dreissena polymorpha*) colonization. *J. Great Lakes Res.* 21:465–75.

Fahnenstiel, G.L, Krause, A. E., McCormick, M. J., Carrick, H. J. & Schelske, C. L. 1998. The structure of the planktonic food web in the St. Lawrence Great Lakes. *J. Great Lakes Res.* 24:531–54.

Fahnenstiel, G. L. & Scavia, D. 1987. Dynamics of Lake Michigan phytoplankton: primary production and growth. *Can. J. Fish Aquat. Sci.* 44:499–508.

Fahnenstiel, G.L, Scavia, D., Lang, G. A., Saylor, J. H., Miller, G. S. & Schwab, D. J. 1988. Impact of inertial period internal waves on fixed-depth primary production estimates. *J. Plankton. Res.* 10:77–87.

Fahnenstiel, G. L., Stone, R. A., McCormick, M. J., Schelske, C. L. & Lohrenz, S. E. 2000. Spring isothermal mixing in the Great Lakes: evidence of nutrient limitation in a sub-optimal light environment. *Can. J. Fish Aquat. Res.* 57:1901–10.

Freedman, P. L. 1974. *Saginaw Bay: An evaluation of existing and historical conditions*. U.S. Environmental Protection Agency Report, EPA-905/9-74-003. Washington, DC.

Frouin, R. & Iacobellis, S. F. 2002. Influence of phytoplankton on the global radiation budget. *J. Geophys. Res.* 107(D19, 4377), doi:10.1029/2001JD000562. refer to: http://www.agu.org/pubs/crossref/2002. . ./2001JD000562.shtml.

Fuller, K., Shear, H. & Wittig, J. [Eds.] 1995. *The Great Lakes: An Environmental Atlas and Resource Book*. 3rd ed. Government of Canada & United States Environmental Protection Agency, Ottawa, Ontario, 46 pp.

Garson, G. D. 1991. Interpreting neural-network connection weights. *Artif. Intell. Expert* 6:47–51.

Gedeon, T. D. 1997. Data mining of inputs: analyzing magnitude and functional measures. *Int. J. Neural Sys.* 8:209–18.

Gurbuz, H., Kirvak, E., Soyupak, S. & Yerli, S. 2003. Predicting dominant phytoplankton quantities in a reservoir by using neural networks. *Hydrobiologia* 504:133–41.

Harris, G. P. 1986. *Phytoplankton Ecology: Structure, Function and Fluctuation*. Chapman & Hall, London, 384 pp.

Harris, G. P. 1994. Pattern, process and prediction in aquatic ecology–a limnological view of some general ecological problems. *Freshwater Biol.* 15:261–66.

Hartig, J. H. & Wallen, D. G. 1984. Seasonal variation of nutrient limitation in western Lake Erie. *J. Great Lakes Res.* 10: 449–60.

Heath, R. T., Fahnenstiel, G. L., Gardner, W. S., Cavaletto, J. F. & Hwang, S. J. 1995. Ecosystem-level effects of zebra mussels (*Dreissena polymorpha*): an enclosure experiment in Saginaw Bay, Lake Huron. *J. Great Lakes Res.* 21:501–16.

Heath, R. T., Sturtevant, R., Shoup, D. & Enflo, P. 2000. Modeling the effects of nutrient concentrations on community production and ecosystem stability: framework for a Great Lakes model. *In* Tulen, L. & dePinto, J. [Eds.] *Great Lakes Modeling Summit: Focus on Lake Erie*. International Joint Commisison, Windsor, ON, pp. 37–50.

Hedgpeth, J. W. 1977. Models and muddles; some philosophical observations. *Helgol. wiss. Meer.* 30:92–104.

Huang, S. H. & Xing, H. 2002. Extract intelligible and concise fuzzy rules from neural networks. *Fuzzy Sets Sys.* 132:233–43.

Jeong, K.-S., Recknagel, F. & Joo, G.-J. 2003. Prediction and elucidation of population dynamics of the blue-green algae *Microcystis aeruginosa* and the diatom *Stephanodiscus hantzschii* in the Nakdong River-Reservoir system (South Korea) by a recurrent artificial neural network. *In* Recknagel, F. [Ed.] *Ecological Informatics; Understanding Ecology by Biologically-Inspired Computation*. Springer-Verlag, Berlin, pp. 195–213.

Johengen, T. H., Nalepa, T. F., Fahnenstiel, G. L. & Goudy, G. 1995. Nutrient changes in Saginaw Bay, Lake Huron, after the establishment of the zebra mussel (*Dreissena polymorpha*) colonization. *J. Great Lakes Res.* 21:465–75.

Johengen, T., Nalepa, T., Lang, G., Fanslow, D., Vanderploeg, H. & Agy, M. 2000. *Physical and Chemical Variables of Saginaw Bay in 1994–1996. National Oceanic & Atmsopheric Administration Technical Memorandum, Great Lakes Environmental Research Laboratory-115*. U. S. Department of Commerce, Washington, DC, 39 pp.

Karul, C., Soyupak, S., Cilesiz, A., Akbay, N. & Germen, E. 2000. Case studies on the use of neural networks in eutrophication modeling. *Ecol. Model.* 134:145–52.

Klarer, D. M. & Millie, D. F. 1994. Regulation of phytoplankton dynamics in a Laurentian Great Lakes estuary. *Hydrobiologia* 286:97–108.

LaBaugh, J. W. 1995. Relation of algal biovolume to chlorophyll *a* in selected lakes and wetlands in the north-central United States. *Can. J. Fish Aquat. Sci.* 52:416–24.

Lee, J. H. W., Huang, Y., Dickman, M. & Jayawardena, A. W. 2003. Neural network modeling of coastal algal blooms. *Ecol. Model.* 159:179–201.

Lohrenz, S. E., Fahnenstiel, G. L., Millie, D. F., Schofield, O. M. E. & Bergmann, T. 2004. Spring phytoplankton photosynthesis, growth, and primary production and relationships to a recurrent coastal sediment plume and river inputs in southeastern Lake Michigan. *J. Geophys. Res.* 109(C10S14), DOI: 10.1029/2004JC002383. refer to: http://www.agu.org/pubs/crossref/2004. . ./2004JC002383.shtml.

Maier, H. R., Dandy, G. C. & Burch, M. D. 1998. Use of artificial neural networks for modeling cyanobacteria *Anabaena* spp. in the River Murray, South Australia. *Ecol. Model.* 105:257–72.

Makarewicz, J. C. & Bertram, P. 1991. Evidence for the restoration of the Lake Erie ecosystem. *Bioscience* 41:216–23.

Mazumder, A. 1994. Patterns of algal biomass in dominant odd- vs. even-link lake ecosystems. *Ecology* 75:1141–49.

Millie, D. F., Dionigi, C. P., Schofield, O. M. E., Kirkpatrick, G. J. & Tester, P. A. 1999. What is the importance for understanding the molecular, cellular, and ecophysiological bases of harmful algal blooms? *J. Phycol.* 35:1353–55.

Millie, D. F., Fahnenstiel, G. L., Carrick, H. J., Lohrenz, S. E. & Schofield, O. M. E. 2002. Phytoplankton pigments in coastal Lake Michigan: distributions during the spring isothermal period and relation with episodic sediment resuspension. *J. Phycol.* 38:639–48.

Millie, D. F., Fahnenstiel, G. L., Lohrenz, S. E., Carrick, H. J., Johengen, T. H. & Schofield, O. M. E. 2003. Physical-biological coupling in southern Lake Michigan: influence of episodic resuspension on phytoplankton. *Aquat. Ecol.* 37:393–408.

Millie, D. F., Paerl, H. W. & Hurley, J. P. 1993. Microalgal pigment assessments using high-performance liquid chromatography: a synopsis of organismal and ecological applications. *Can. J. Fish. Aquat. Sci.* 50:2513–27.

Millie, D. F., Weckman, G. R., Paerl, H. W., Pinckney, J. L., Bendis, B. J., Pigg, R. J. & Fahnenstiel, G. L. 2006. Neural network modeling of estuarine indicators: hindcasting phytoplankton biomass and net ecosystem production in the Neuse (North Carolina) and Trout (Florida) Rivers. *Ecol. Ind.* doi: 10.1016/j.ecoind.2005.08.021.

Milne, L. K. 1995. Feature selection with neural networks with contribution measures. *In* Yao, X. [Ed.] *Proceedings of the Eighth Australian Joint Conference on Artificial Intelligence, AI'95*. World Scientific Publ. Co., Singapore, pp. 215–21.

Moersdorf, P. & Meindl, E. 2003. Sustained Ocean Observations and the Role of NOAA's Marine Observation Network. *Mar. Sci. Technol. J.* 37:67–74.

Murray, A. G. & Parslow, J. S. 1999. The analysis of alternative formulations in a simple model of a coastal ecosystem. *Ecol. Model.* 119:146–66.

Nalepa, T. F. & Fahnenstiel, G. L. [Ed.] 1995. Zebra Mussels in the Saginaw Bay, Lake Huron Ecosystem. *J. Great Lakes Res.* 21:411–573.

Nalepa, T. F., Fahnenstiel, G. L. & Johengen, T. H. 1999. Impacts of the zebra mussel (*Dreissena polymorpha*) on water quality: A case study in Saginaw Bay, Lake Huron. *In* Claudi, R. & Leach, J. H. [Eds.] *Non-Indigenous Freshwater Organisms in North America: Their Biology and Impact*. CRC Press, Lewis Publishers, Boca Raton, FL, pp. 255–71.

Nalepa, T., Fahnenstiel, G., McCormick, M., Johengen, T., Lang, G., Cavaletto, J. & Goudy, G. 1996. *Physical and Chemical Variables of Saginaw Bay in 1991–1993*. National Oceanic & Atmospheric Administration Technical Memorandum, Great Lakes Environmental Research Laboratory-91. U. S. Department of Commerce, Washington, DC, 57 pp.

Nalepa, T. F., Fanslow, D. L., Lansing, M. B., Lang, G. A., Ford, M., Gostenik, G. & Hartson, D. J. 2002. *Abundance, Biomass, and Species Composition of Benthic Macroinvertebrate Populations in Saginaw Bay, Lake Huron, 1987–1996*. National Oceanic & Atmospheric Administration Technical Memorandum, Great Lakes Environmental Research Laboratory-122. U. S. Department of Commerce, Washington, DC, 27 pp.

Nalepa, T. F., Wojcik, J. A., Fanslow, D. L. & Lang, G. A. 1995. Initial colonization of the zebra mussel (*Dreissena polymorpha*) in Saginaw Bay, Lake Huron: population recruitment, density, and size structure. *J. Great Lakes Res.* 21:417–34.

Olden, J. D. 2000. An artificial neural network approach for studying phytoplankton succession. *Hydrobiologia* 436: 131–43.

Olden, J. D. & Jackson, D. A. 2002. Illuminating the "black box": understanding variable contributions in artificial neural networks. *Ecol. Model.* 154:135–50.

Özesmi, S. L. & Özesmi, U. 1999. An artificial neural network approach to spatial habitat modelling with interspecific interaction. 116:15–31.

Paerl, H. W. 1988. Nuisance phytoplankton blooms in coastal, estuarine, and inland waters. *Limnol. Oceanogr.* 33:823–47.

Paerl, H. W., Dyble, J., Pinckney, J. L., Valdes, L. M., Millie, D. F., Moisander, P. H., Morris, J. T., Bendis, B. & Piehler, M. F. 2005. Using microalgal indicators to assess human and climactically-induced ecological change in estuaries. *In* Bortone, S. A. [Ed.] *Estuarine Indicators*. CRC Press, Boca Raton, FL, pp. 145–74.

Paerl, H. W., Valdes, L. M., Pinckney, J. L., Piehler, M. F., Dyble, J. & Moisander, P. H. 2003. Phytoplankton photopigments as indicators of estuarine and coastal eutrophication. *BioScience* 53:953–64.

Principe, J. C., Euliano, N. R. & Lefebrve, W. C. 2000. *Neural and Adaptive Systems: Fundamentals through Simulation*. John Wiley & Sons Inc., New York, 656 pp.

Reckhow, K. H. 1994. Water quality simulation modeling and uncertainty analysis for risk assessment and decision making. *Ecol. Model.* 72:1–20.

Recknagel, F. 2003a. *Ecological Informatics; Understanding Ecology by Biologically Inspired Computation*. Springer-Verlag, Berlin, 398 pp.

Recknagel, F. 2003b. Ecological applications of adaptive agents. *In* Recknagel, F. [Ed.] *Ecological Informatics; Understanding Ecology by Biologically-Inspired Computation*. Springer-Verlag, Berlin, pp. 73–88.

Scavia, D. & Fahnenstiel, G. L. 1987. Dynamics of Lake Michigan's phytoplankton: mechanisms controlling epilimnetic communities. *J. Great Lakes Res.* 13:103–20.

Scavia, D., Powers, W. F., Canale, R. P. & Moody, J. L. 1981. Comparison of first-order error analysis and Monte Carlo simulation in time-dependent lake eutrophication models. *Water Resour. Res.* 17:1051–59.

Schelske, C. L., Stoermer, E. F., Fahnenstiel, G. L. & Haibach, M. 1986. Phosphorus enrichment, silica utilization, and biogeochemical silica depletion in the Great Lakes. *Can. J. Fish Aquat. Sci.* 43:407–15.

Schelske, C. L. 1979. The role of phosphorus in Great Lakes eutrophication: is there a controversy? *J. Fish Res. Board Can.* 36:286–88.

Schelske, C. L., Davis, C. O. & Feldt, L. E. 1984. Growth responses of river and lake phytoplankton populations in Lake Michigan water. *Verh Internat Verein Limnol.* 22:445–51.

Schelske, C. L., Feldt, L. E., Simmons, M. S. & Stoermer, E. F. 1974. Storm induced relationships among chemical conditions and phytoplankton in Saginaw Bay and western Lake Huron. *In* International Association for Great Lakes Research. *Proceedings of the 17th Conference on Great Lakes Research*. Ann Arbor, Michigan, MI, pp 78–91.

Schleiter, I. M., Obach, M., Wagner, R., Werner, H., Schmidt, H.-H. & Borchardt, D. 2003. *In* Recknagel, F. [Ed.] *Ecological Informatics; Understanding Ecology by Biologically Inspired Computation*. Springer-Verlag, Berlin, pp. 109–26.

Schofield, O. & Tivey, M. 2004. Building a window to the sea: Ocean Research Interactive Observing Networks (ORION). *Oceanography* 17:105–11.

Schwab, D. J. & Bedford, K. W. 1999. The Great Lakes Forecasting System. *In* Mooers, C. N. K. [Ed.] *Coastal Ocean Prediction, Coastal and Estuarine Studies 56*. American Geophysical Union, Washington, DC, pp. 157–73.

Schwab, D. J., Leshkevich, G. A. & Muhr, G. C. 1992. Satellite measurements of surface water temperature in the Great Lakes: Great Lakes Coastwatch. *J. Great Lakes Res.* 18: 247–58.

Smith, A. S. & Mason, A. K. 1997. Cost estimation predictive modeling: regression versus neural network. *Eng. Econ.* 42:137–61.

Snedecor, G. W. & Cochran, W. G. 1980. *Statistical Methods*. 7th ed. Iowa State University Press, Ames, 507 pp.

Strutton, P. G., Mitchell, J. G. & Parslow, J. S. 1996. Non-linear analysis of chlorophyll *a* transects as a method of quantifying spatial structure. *J. Plankton Res.* 18:1717–26.

Sugihara, G., Grenfell, B. & May, R. M. 1990. Distinguishing error from chaos in ecological time series. *Philos. Trans. R. Soc. London B* 330:235–51.

SYSTAT 10. 2000. *Statistics 1*. SPSS Inc., Chicago, IL, 663 pp.

Tester, P. A., Varnum, S. M., Culver, M. E., Eslinger, D. L., Stumpf, R. P., Swift, R. N., Yungel, J. K., Black, M. D. & Litaker, R. W. 2003. Airborne detection of ecosystem responses to an extreme event: phytoplankton displacement and abundance after hurricane induced flooding in the Pamlico-Albemarle Sound System, North Carolina. *Estuaries* 26:1353–64.

Vanderploeg, H. A., Liebig, J. R., Carmichael, W. W., Agy, M. A., Johengen, T. H., Fahnenstiel, G. L. & Nalepa, T. F. 2001. Zebra mussel (*Dreissena polymorpha*) selective filtration promoted toxic *Microcystis* blooms in Saginaw Bay (Lake Huron) and Lake Erie. *Can. J. Fish. Aquat. Sci.* 58:1208–21.

Walsh, J. J., Weisberg, R. H., Dieterle, D. A., He, R., Darrow, B. P., Jolieff, J. K., Lester, K. M., Vargo, G. A., Kirkpatrick, G. J., Fanning, K. A., Sutton, T. T., Jochens, A. E., Biggs, D. C., Nababan, B., Hu, C. & Muller-Karger, F. E. 2003. Phytoplankton response to intrusions of slope water on the West Florida Shelf: models and observations *J. Geophys. Res.*, 108 (C6), 3190, doi: 10.1029/2002JC001406.

Walsh, J. J., Penta, B., Dieterle, D. A. & Bissett, W. P. 2001. Predictive ecological modeling of harmful algal blooms. *Hum. Ecol. Risk Asses.* 7:1369–83.

Wigham, P. A. & Fogel, G. B. 2003. Ecological applications of evolutionary computation. *In* Recknagel, F. [Ed.] *Ecological Informatics; Understanding Ecology by Biologically-Inspired Computation*. Springer-Verlag, Berlin, pp. 49–71.